



# An overlook on statistical inference issues for stochastic damping Hamiltonian systems under the fluctuation-dissipation condition

Patrick Cattiaux, Jose Rafael León, Angie Pineda Centeno, Clémentine Prieur

## ► To cite this version:

Patrick Cattiaux, Jose Rafael León, Angie Pineda Centeno, Clémentine Prieur. An overlook on statistical inference issues for stochastic damping Hamiltonian systems under the fluctuation-dissipation condition. *Statistics*, 2017, Special Issue in Honour of Paul Doukhan: Selected Papers from the Conference on the Occasion of Paul Doukhan's 60th Birthday: Dependence, Limit Theorems and Applications, 51 (1), pp.11-29. 10.1080/02331888.2016.1259807 . hal-01405427

**HAL Id: hal-01405427**

**<https://hal.science/hal-01405427>**

Submitted on 8 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

To appear in *Statistics: A Journal of Theoretical and Applied Statistics*  
Vol. 00, No. 00, Month 20XX, 1–22

## ***An overlook on statistical inference issues for stochastic damping hamiltonian systems under the fluctuation-dissipation condition***

P. Cattiaux<sup>a</sup>, J. R. León<sup>b,d</sup>, A. A. Pineda Centeno<sup>b</sup> and C. Prieur<sup>c,d\*</sup>

<sup>a</sup>*Univ. de Toulouse, CNRS, Institut de mathématiques de Toulouse, France;*

<sup>b</sup>*Escuela de Matemática, Facultad de Ciencias, Universidad Central de Venezuela, Caracas;*

<sup>c</sup>*Univ. Grenoble Alpes, CNRS, LJK, F-38000 Grenoble, France;* <sup>d</sup>*Inria project/team AIRSEA*

(Received 00 Month 20XX; final version received 00 Month 20XX)

In this paper, we are interested in nonparametric inference issues for stochastic damping hamiltonian systems under the fluctuation-dissipation condition. This condition relates the magnitude of the dissipative term and the magnitude of the random term. The precise balance between the drift term which removes energy in average and the stochastic term provided by the fluctuation-dissipation relation insures that the canonical measure is preserved by the dynamics. In this framework, it is possible to give an explicit construction of a Lyapunov function and thus to prove exponential ergodicity. Then, we consider various estimation procedures and provide also a numerical section, where simulations are conducted.

**Keywords:** hypoelliptic diffusion; variance estimation; fluctuation-dissipation models

*AMS Subject Classification:* Primary 62M05; secondary 60H10; 60F05; 35H10

### **1. Introduction**

The stochastic damping hamiltonian system is defined as the following system of SDE

$$\begin{cases} dX_t = Y_t dt \\ dY_t = \sigma(X_t, Y_t) dW_t - (c(X_t, Y_t)Y_t + \nabla V(X_t))dt. \end{cases} \quad (1.1)$$

We denote the vector solution as  $Z_t = (X_t, Y_t) \in \mathbb{R}^{2d}$ . The matrix  $c$  is called the damping force and  $V$  the potential, the symmetric matrix  $\sigma(x, y)$  is the diffusion term and  $W$  is a standard Brownian motion defined in  $\mathbb{R}^d$ .

In the present paper, we will study some inference issues in the following particular case:

$$\begin{cases} dX_t = Y_t dt \\ dY_t = (2\beta^{-1})^{1/2} \sigma(X_t) dW_t - (\sigma^2(X_t)Y_t + \nabla V(X_t))dt. \end{cases} \quad (1.2)$$

Observe that the diffusion term in (1.2),  $\tilde{\sigma}(x) = (2\beta^{-1})^{1/2} \sigma(x)$ , depends only on the  $x$  coordinate and on an unknown parameter  $\beta$ . Moreover, the damping force has the form  $c(x, y) = \frac{\beta}{2} \tilde{\sigma}^2(x)$ . Since Einstein, this last relationship between the damping force

---

\*Corresponding author. Email: clementine.prieur@imag.fr

and the diffusion term is known as the fluctuation-dissipation relation and has numerous applications. For instance, it appears as a tool for the simulation of molecular dynamics (see [7] and references therein and also the recent book by Lelièvre *et al.* [8, Section 2.2.3]). System (1.2) also appears as limit of the Ehrenfest nuclei dynamics and is called Langevin dynamics (see [11]).

Under additional assumptions (see Section 2 for details), we can prove that the Markov process solution of (1.2) has an invariant measure whose density is proportional to the Boltzmann distribution  $\exp(-\beta H(x, y))$  where  $H(x, y) = \frac{1}{2}|y|^2 + V(x)$ .

In this work, by following the spirit of our recent papers [1–4], we aim to propose convergent estimators of the various unknown quantities involved in the formulation (1.2) of fluctuation-dissipation stochastic damping hamiltonian systems. The fluctuation-dissipation relation allows to simplify the assumptions ensuring the ergodicity of the process solution of (1.2). These assumptions are stated in Section 2. In Section 3, we detail two classes of examples. We first provide an explicit construction of the Lyapunov function in the case of hypoelliptic diffusions on  $\mathbb{R}^{2d}$  under a drift condition. We then consider the class of periodic potentials defined on the torus. In Section 4, we state our results concerning the inference of such models. Section 5 is devoted to numerical experiments.

## 2. Hypotheses and notation

In this section, we precise the context of our study. We will state some hypotheses implying the existence and uniqueness of the solution of (1.2). These hypothesis also imply the existence and unicity of an invariant probability measure with smooth density with respect to the Lebesgue measure. Finally, an additional assumption is provided to ensure the exponential ergodicity of the underlying process, which is crucial for establishing our estimation results. Before stating the hypothesis let us introduce some notation. We will denote by  $C^\infty(A, B)$  the infinitely differentiable functions  $f : A \rightarrow B$  where  $A$  and  $B$  are finite dimensional normed spaces if  $B = \mathbb{R}$  we denote simply  $C^\infty(A)$ . The symbol  $\mathcal{M}(\mathbb{R}^m, \mathbb{R}^m)$  denotes the normed space of square matrices.

We assume that our system satisfies the following hypotheses.

- $\mathcal{H}_0$  the diffusion matrix  $\sigma$  is symmetric,  $C^\infty(\mathbb{R}^d, \mathcal{M}(\mathbb{R}^d, \mathbb{R}^d))$ , bounded as well as its first and second partial derivatives and uniformly elliptic, i.e.  $\forall x \in \mathbb{R}^d$ ,  $\sigma(x) \geq \sigma_0 Id$  (in the sense of quadratic forms) for a positive constant  $\sigma_0 > 0$ ;
- $\mathcal{H}_1$  the potential  $V$  is lower bounded and  $C^\infty(\mathbb{R}^d)$ .

*Remark 2.1* Let  $\sigma^*$  denote the transposed matrix of  $\sigma$ . Actually the law of the process solution of (1.2) depends on  $\sigma\sigma^*$  (which is the second order term of the infinitesimal generator). If this symmetric matrix is smooth, it is well known that one can find a smooth symmetric square root of it, which is the choice we make for  $\sigma$ , hence  $\sigma\sigma^* = \sigma^2$ . In the following, we use the notation  $\sigma(x) = (\sigma_{ij}(x))_{1 \leq i, j \leq d}$ .

*Remark 2.2* Given the particular form of the damping term, we have that  $c(x, y)$  is a symmetric and uniformly elliptic matrix.

Let  $\Omega = C(\mathbb{R}^+, \mathbb{R}^{2d})$  denote the space of continuous functions from  $\mathbb{R}^+$  to  $\mathbb{R}^{2d}$ , equipped with the usual compact convergence topology. According to Wu [13, Lemma 1.1], Hypotheses  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  and Remark 2.2 entail that for every initial state  $z = (x, y) \in \mathbb{R}^{2d}$ , the stochastic differential equation (1.2) admits a unique weak solution  $\mathbf{P}_z$  on  $\Omega$ , and that this solution is non-explosive. Let  $(P_t(z, dz'), t > 0)$  be the semigroup of transition

probability kernels of the Markov process  $((Z_t)_{t>0}, (\mathbf{P}_z)_{z \in \mathbb{R}^{2d}})$  solution of (1.2). Let us prove that  $P_t$  is strong Feller for each  $t > 0$ . Let  $C^{1,2}(\mathbb{R}^{2d})$  be the space of all functions  $f(x, y)$  such that  $\partial_x^k f$ ,  $k = 0, 1$  and  $\partial_y^l f$ ,  $l = 1, 2$  are continuous on  $\mathbb{R}^{2d}$ . The infinitesimal generator  $\mathcal{L}$  of (1.2) is given by: for any  $f \in C^{1,2}(\mathbb{R}^{2d})$ ,

$$\mathcal{L}f(x, y) = \left( \frac{1}{\beta} \sum_{i,j=1}^d (\sigma^2)_{ij}(x) \partial_{y_i} \partial_{y_j} + y \cdot \nabla_x - (\sigma^2(x)y + \nabla V(x)) \cdot \nabla_y \right) f(x, y). \quad (2.3)$$

Let us define the differential operators  $A_k$ ,  $k = 0, \dots, d$  as

$$\begin{aligned} A_0 &= \sum_{j=1}^d y_j \cdot \partial_{x_j} - \sum_{j=1}^d \partial_{x_j} V(x) \partial_{y_j} - \sum_{j=1}^d y_j \sum_{l=1}^d (\sigma^2)_{jl}(x) \partial_{y_l} \\ A_k &= \sum_{j=1}^d \sigma_{kj}(x) \partial_{y_j} \quad k = 1, \dots, d. \end{aligned}$$

We can decompose  $\mathcal{L}$  in the Hörmander form, i.e.  $\mathcal{L} = \frac{1}{\beta} \sum_{k=1}^d A_k^2 + A_0$ . Let  $[A, B]$  denote the Lie bracket between the operators  $A$  and  $B$ . As  $\sigma^2$  is uniformly elliptic, it is possible to prove that (see [8] pages 91-92 for more details):

$$\text{Span}(A_0, \dots, A_d, [A_0, A_1], \dots, [A_0, A_d]) = \text{Span}(\partial_{x_1}, \dots, \partial_{x_d}, \partial_{y_1}, \dots, \partial_{y_d}).$$

Then, according to Hörmander's theorem on hypoellipticity, we deduce that  $P_t$  is strong Feller. Moreover, if a stationary probability distribution  $\mu$  having a positive density with respect to the Lebesgue measure exists, then the irreducibility property is verified, the stationary probability distribution is unique, and pathwise ergodicity holds for any initial condition  $z \in \mathbb{R}^{2d}$  (see [8, Theorem 2.7] and references therein). Let us now consider the probability measure

$$d\mu(x, y) = p_s^\beta(x, y) dx dy \text{ with } p_s^\beta(x, y) = C(\beta) e^{-\beta H(x, y)} > 0. \quad (2.4)$$

It is possible to prove that  $\mu$  is a stationary probability distribution. Indeed, for any smooth test function  $\varphi$ :

$$\begin{aligned} & \int_{\mathbb{R}^{2d}} \mathcal{L}\varphi(x, y) p_s^\beta(x, y) dx dy \\ &= \frac{1}{\beta} \int_{\mathbb{R}^{2d}} (\nabla_y p_s^\beta \nabla_x \varphi - \nabla_x p_s^\beta \nabla_y \varphi) dx dy + \frac{1}{\beta} \int_{\mathbb{R}^{2d}} \text{div}_y(p_s^\beta \sigma^2(x) \nabla_y \varphi) dx dy = 0. \end{aligned}$$

Thus  $\mathcal{L}^*(\mu) = 0$  and we conclude that the probability measure  $\mu$  is an invariant measure. We refer to [8] page 90 for more details on the above calculation.

In the following, we reinforce Hypotheses  $\mathcal{H}_0$  and  $\mathcal{H}_1$  by assuming that the stationary solution of (1.2) is exponentially  $\beta$ -mixing (see Hypothesis  $\mathcal{H}_2$  below).

- $\mathcal{H}_2$  there exist two constants  $C > 0$ ,  $\lambda > 0$  such that for all  $t > 0$ ,

$$\int_{\mathbb{R}^{2d}} \sup_{f \text{ measurable, } |f| \leq 1} \left| P_t f(z) - \int_{\mathbb{R}^{2d}} f d\mu \right| \mu(dz) \leq C e^{-\lambda t}. \quad (2.5)$$

Assumption  $\mathcal{H}_2$  will allow the use of adequate covariance inequalities for the proof of inference results stated in this paper (see [5] and references therein).

In the following we present two classes of examples for which the exponential ergodicity is satisfied, by an explicit construction of the Lyapunov function.

### 3. Two important classes of examples

In this section, we consider two classes of examples for which exponential ergodicity is proved.

#### 3.1. Hypoelliptic diffusion on $\mathbb{R}^{2d}$ under a drift condition

Let us consider System (1.2) with Assumptions  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . We know from Section 2 that the process  $Z$  solution of (1.2) has a unique invariant probability measure described by (2.4). The infinitesimal generator  $\mathcal{L}$  of System (1.2) is given by (2.3). In the sequel we assume that the potential satisfies condition (0.5) of [13]:

$$\langle \nabla V(x), \frac{x}{\|x\|} \rangle \rightarrow \infty \text{ as } \|x\| \rightarrow \infty \text{ with } \|\cdot\| \text{ the usual euclidean norm on } \mathbb{R}^d. \quad (3.6)$$

From [13, Theorem 4.1], we could then directly conclude that exponential convergence holds. However, from a pedagogical point of view, we think it is worth detailing the different steps leading to the construction of an appropriate Lyapunov test function. The construction of the Lyapunov function we detail below follows the main lines of the proof of Theorem 4.1 in [13], but is slightly different. For sake of simplicity in the notation, we assume without loss of generality that  $\beta = 2$  in the sequel. We first introduce a function  $G : \mathbb{R}^d \rightarrow \mathbb{R}^d$  defined as:

$$G(x) = \Phi(\|x\|) \frac{x}{\|x\|} \text{ for } x \neq \mathbf{0} \in \mathbb{R}^d, \quad G(\mathbf{0}) = 0,$$

with  $\Phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  a non decreasing smooth function equal to zero on a small neighborhood of 0 and equal to 1 for  $\|x\| \geq 1$ . By construction we have  $G \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$  and  $\lim_{\|x\| \rightarrow \infty} \|\partial_{x_j} G_i(x)\|_{HS} = 0$ , where  $\|\cdot\|_{HS}$  denotes the Hilbert-Schmidt norm of operators. We can now propose the following form for the Lyapunov function:

$$\Psi(x, y) = e^{F(x, y) - \inf_{\mathbb{R}^d} F} \geq 1,$$

where

$$F(x, y) = a\left(\frac{1}{2}\|y\|^2 + V(x)\right) + b < G(x), y >.$$

The constants  $a$  and  $b$  will be calibrated later on. Following the proof of [13, Theorem 3.1], we can prove that:

$$-\frac{\mathcal{L}\Psi}{\Psi} = -\mathcal{L}F - \frac{1}{2}\|\sigma(x)\nabla_y F\|^2.$$

Then, as

$$\begin{aligned}\mathcal{L}\left(a\left(\frac{1}{2}\|y\|^2\right)\right) &= a\left(\frac{1}{2}\sum_{i=1}^d(\sigma^2)_{ii}(x)-\langle\sigma^2(x)y,y\rangle-\langle y,\nabla V(x)\rangle\right) \\ \mathcal{L}(aV(x)) &= a\langle\nabla V(x),y\rangle \\ \mathcal{L}(b\langle G(x),y\rangle) &= b\sum_{j=1}^d\sum_{i=1}^d(\partial_{x_j}G_i(x))y_iy_j-b\langle y,\sigma^2(x)G(x)\rangle-b\langle\nabla V(x),G(x)\rangle\end{aligned}$$

we get

$$\begin{aligned}\mathcal{L}(F(x,y)) &= a\left(\frac{1}{2}\sum_{i=1}^d(\sigma^2)_{ii}(x)-\langle\sigma^2(x)y,y\rangle\right) \\ &\quad +b\left(\sum_{j=1}^d\sum_{i=1}^d\partial_{x_j}G_i(x)y_iy_j-\langle y,\sigma^2(x)G(x)\rangle-\langle\nabla V(x),G(x)\rangle\right).\end{aligned}$$

Moreover

$$\frac{1}{2}\|\sigma(x)\nabla_y F\|^2 = \frac{1}{2}\|\sigma(x)(ay+bG(x))\|^2.$$

Hence

$$\begin{aligned}-\frac{\mathcal{L}\Psi}{\Psi} &= -a\left(\frac{1}{2}\sum_{i=1}^d(\sigma^2)_{ii}(x)-\langle\sigma^2(x)y,y\rangle\right) \\ &\quad -b\left(\sum_{j=1}^d\sum_{i=1}^d\partial_{x_j}G_i(x)y_iy_j-\langle y,\sigma^2(x)G(x)\rangle-\langle\nabla V(x),G(x)\rangle\right)-\frac{1}{2}\|\sigma(x)(ay+bG(x))\|^2.\end{aligned}$$

The aim is now to apply [6, Theorem 5.2c] (see also [13, Theorem 2.4]). We already know from Section 2 that we have the strong Feller property and that there exists a unique invariant probability measure whose density  $p_s^\beta(z)$  is positive on  $\mathbb{R}^d$ . By construction we have  $\Psi \geq 1$ , and from Hypotheses  $\mathcal{H}_0$  and  $\mathcal{H}_1$  we know that  $\Psi$  is smooth. Thus, to prove Hypothesis  $\mathcal{H}_2$ , it remains to prove the existence of some compact subset  $\mathbf{K} \subset \mathbb{R}^{2d}$  and of some constants  $\varepsilon, C > 0$  such that

$$-\frac{\mathcal{L}\Psi}{\Psi} \geq \varepsilon \mathbf{1}_{\mathbf{K}^c} - C \mathbf{1}_{\mathbf{K}}. \quad (3.7)$$

*Proof of Inequality (3.7):* From Hypothesis  $\mathcal{H}_0$  we know that there exists two positive constants  $\sigma_0$  and  $\sigma_1$  such that  $\forall x \in \mathbb{R}^d, \sigma_0 Id \leq \sigma(x) \leq \sigma_1 Id$  in the sense of quadratic forms. Remark that the right hand side of Inequality (3.7) is composed with two terms, the first one being concerned with what happens on the non bounded domain  $\mathbf{K}^c$  and the second one with what happens on the compact set  $\mathbf{K}$ .

Let us first write the following bound: for all  $x \in \mathbb{R}^d$ ,

$$\|\sigma(x)(ay+bG(x))\|^2 \leq \sigma_1^2(2a^2\|y\|^2+2b^2\|G(x)\|^2) = \sigma_1^2(2a^2\|y\|^2+2b^2). \quad (3.8)$$

We then split the proof in two parts.

**Step 1:**  $x \in \mathbb{R}^d$ ,  $\|y\| \leq L'$ . Let us first prove that there exists some positive constant  $\varepsilon'$  such that:

$$\sup_{x \in \mathbb{R}^d} (a\sigma^2(x) - b\partial_{x_j} G_i(x)) \geq (\sigma_0^2 a^2 - \varepsilon') Id \text{ for all } y \in \mathbb{R}^d.$$

The above inequality is to be understood in the sense of quadratic forms. We have:

$$\langle (a\sigma^2(x) - b\partial_{x_j} G_i(x))y, y \rangle \geq \sigma_0^2 a \|y\|^2 - b \langle \partial_{x_j} G_i(x)y, y \rangle.$$

We have:

$$\lim_{\|x\| \rightarrow \infty} \|\partial_{x_j} G_i(x)\|_{HS} = 0. \quad (3.9)$$

We can thus choose  $b$  such that

$$b \sup_{x \in \mathbb{R}^d} \|\partial_{x_j} G_i(x)\|_{HS} \leq \varepsilon'.$$

We then get:

$$\sup_{x \in \mathbb{R}^d} \langle (a\sigma^2(x) - b\partial_{x_j} G_i(x))y, y \rangle \geq (a\sigma_0^2 - \varepsilon') \|y\|^2. \quad (3.10)$$

Then, from the definition of  $G$  and from Assumption (3.6) we get the existence of some positive constants  $M$ ,  $M'$  such that:

$$\sup_{x \in \mathbb{R}^d} \langle \nabla V(x), G(x) \rangle \geq -M \quad (3.11)$$

$$\text{and } \sup_{x \in \mathbb{R}^d} (b \langle y, \sigma^2(x) G(x) \rangle) \geq -M' \|y\|. \quad (3.12)$$

Moreover we have: for all  $x \in \mathbb{R}^d$ ,

$$-a \frac{1}{2} \sum_{i=1}^d (\sigma^2)_{ii}(x) \geq -a \frac{d}{2} \sigma_1^2, \quad a \langle \sigma^2(x)y, y \rangle \geq a\sigma_0^2 \|y\|^2. \quad (3.13)$$

Then, from (3.8), (3.10), (3.17), (3.12) and (3.13), we get, for all  $x, y \in \mathbb{R}^d$ :

$$-\frac{\mathcal{L}\Psi}{\Psi} \geq (a\sigma_0^2 - a^2\sigma_1^2 - \varepsilon' + Mb) \|y\|^2 - M' \|y\| - \left( \frac{ad\sigma_1^2}{2} + b^2\sigma_1^2 \right). \quad (3.14)$$

Let us choose  $a$  such that  $a\sigma_0^2 - a^2\sigma_1^2 > 0$ . We say that  $a$  satisfies Condition  $\mathcal{C}_1$ . Then, we choose  $b$  such that  $\varepsilon'$  satisfies  $a\sigma_0^2 - a^2\sigma_1^2 - \varepsilon' > 0$ . We say that  $b$  satisfies Condition  $\mathcal{C}_2$ . We then get: for all  $L' > 0$ , for all  $x \in \mathbb{R}^d$  and for all  $\|y\| \leq L'$ ,

$$-\frac{\mathcal{L}\Psi}{\Psi} \geq -M' \|y\| - \left( \frac{ad\sigma_1^2}{2} + b^2\sigma_1^2 \right) \geq -C(L') \text{ with } C(L') = M'L + \left( \frac{ad\sigma_1^2}{2} + b^2\sigma_1^2 \right) > 0.$$

**Step 2:**  $x \in \mathbb{R}^d$ ,  $\|y\| > L'_0$ . Let  $\varepsilon > 0$ . As  $a$  satisfies  $\mathcal{C}_1$ , we deduce from (3.14) that there exists  $L'_0 > 0$  such that for all  $x \in \mathbb{R}^d$  and for all  $\|y\| > L'_0$ :

$$-\frac{\mathcal{L}\Psi}{\Psi} \geq \varepsilon. \quad (3.15)$$

**Step 3:**  $\|x\| > L_0$ ,  $\|y\| \leq L'_0$ . From (3.8), (3.10), (3.12) and (3.13), we get, for all  $x, y \in \mathbb{R}^d$ :

$$-\frac{\mathcal{L}\Psi}{\Psi} \geq (a\sigma_0^2 - a^2\sigma_1^2)\|y\|^2 + b < \nabla V(x), G(x) > -M'\|y\| - \left( \frac{ad\sigma_1^2}{2} + b^2\sigma_1^2 \right). \quad (3.16)$$

Then, from the definition of  $G$  and from Assumption (3.6) we get:

$$< \nabla V(x), G(x) > \rightarrow \infty \text{ as } \|x\| \rightarrow \infty. \quad (3.17)$$

Thus there exists  $L_0 > 0$  such that for all  $\|y\| \leq L'_0$  and for all  $\|x\| > L_0$  we get (3.15). The compact subset is thus defined as  $\mathbf{K} = \{z = (x, y), \|x\| \leq L_0 \text{ and } \|y\| \leq L'_0\}$ .

**Conclusion:** Under Assumptions  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  and (3.6), it is possible to construct explicitly a Lyapunov function satisfying the minoration (3.7): we take  $\Psi(x, y) = e^{F(x, y) - \inf_{\mathbb{R}^d} F} \geq 1$ , with  $F(x, y) = a(\frac{1}{2}\|y\|^2 + V(x)) + b < G(x), y >$  and with  $a$  and  $b$  satisfying Conditions  $\mathcal{C}_1$  and  $\mathcal{C}_2$ .

### 3.2. Particle in a periodic potential

We now consider  $z = (x, y) \in \mathbb{T}^d \times \mathbb{R}^d$ , where  $\mathbb{T}^d$  denotes the  $d$ -dimensional torus. Under Assumptions  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  and if  $V \geq 1$ , it is possible to construct explicitly a Lyapunov function (see [9, Section 3]) satisfying the lower bound (3.7): we take  $\Psi(x, y) = \frac{\|y\|^2}{2} + V(x)$ .

## 4. Main inference results

The aim of this section is to consider nonparametric inference issues for System (1.2). In the following, we only consider the framework with complete observations, that is the framework where both coordinates  $X$  and  $Y$  are observed. However, most of our results can be extended to the case where only the first coordinate  $X$  is observed.

In [1], a central limit theorem has been proved for the estimation of the stationary density  $p_s^\beta(x, y)$  described by (2.4) in the case where the diffusion term  $\sigma$  is constant. However, as the main ingredient for the proof is the exponential ergodicity, which is here ensured by Assumptions  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , Corollary 4.1 below is straightforward.

Let  $K$  be some  $C^2$  function with compact support  $A$  such that  $\int_A K(x, y) dx dy = 1$ . We may also assume, without loss of generality that  $A$  is a bounded ball. Moreover, we assume that there exists  $m \in \mathbb{N}^*$  such that for all non constant polynomial  $P(x, y)$  with degree less than or equal to  $m$ ,  $\int P(u, v) K(u, v) du dv = 0$ .



We then define

$$\tilde{p}_s^\beta(x, y) := \frac{1}{nb_{1,n}^d b_{2,n}^d} \sum_{i=l_n+1}^{l_n+n} K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right). \quad (4.18)$$

**COROLLARY 4.1** *Assume Hypotheses  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are fulfilled. Assume that the bandwidths  $b_{1,n}$ ,  $b_{2,n}$  and the discretization step  $h_n$  satisfy the following assumptions : (i)  $b_{1,n}$ ,  $b_{2,n}$  and  $h_n \rightarrow 0$ , (ii)  $nb_{1,n}^d b_{2,n}^d \rightarrow +\infty$ , (iii)  $\frac{b_{1,n} b_{2,n}}{h_n^2} \rightarrow 0$ , (iv)  $\exists m \in \mathbb{N}^*$  such that  $nb_{1,n}^d b_{2,n}^d \max(b_{1,n}, b_{2,n})^{2(m+1)} \rightarrow 0$ , (v)  $l_n h_n \rightarrow +\infty$ .*

*Then, starting from any initial point  $z_0 = (x_0, y_0)$ ,*

$$\sqrt{nb_{1,n}^d b_{2,n}^d} \left( \tilde{p}_s^\beta(x, y) - p_s^\beta(x, y) \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(0, p_s^\beta(x, y) \int K^2(s, t) ds dt \mathcal{I}\right).$$

*Proof.* The exponential ergodicity is a consequence of Hypotheses  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  and  $\mathcal{H}_2$ . Theorem 2.10 and Corollary 2.12 in [1] still hold true if  $\sigma$  depends on  $x$  (see Theorem 2.1 in [10]). Thus the result is a straightforward corollary of Theorem 3.3 and Theorem 6.2 in [1].  $\blacksquare$

**Remark 4.2** Further in the paper, we only need the convergence in probability of  $\tilde{p}_s^\beta(x, y)$  to apply Slutsky's Theorem. From covariance inequalities stated in [12] for  $\beta$ -mixing sequences, we know that

$$\sqrt{nh_n b_{1,n}^d b_{2,n}^d} \left( \tilde{p}_s^\beta(x, y) - p_s^\beta(x, y) \right) = O_{\mathbb{P}}(1) \quad (4.19)$$

with  $O_{\mathbb{P}}(1)$  means *stochastically bounded*, and as soon as (i) – (ii), (iv) – (v) in Corollary 4.1 are satisfied.

Let us now focus our attention on the diffusion term  $\sigma(x)$ . We know from Theorem 4.2 in [4] that under Hypotheses  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , it is possible to prove a central limit theorem for the estimation of  $\mathbb{E}_\mu \sigma^2(X_0, Y_0)$  where  $\mu(dz) = p_s^\beta(z) dz$  is the unique stationary density for System (1.2) described by (2.4). Indeed, let  $g(x, y) = -(\sigma^2(x)y + \nabla V(x))$ . We know from the expression of  $p_s^\beta(z)$  that  $\mathbb{E}_\mu(|g(Z_0)|^r) < \infty$  for all  $r > 0$ . It yields the first part of Assumption  $\mathcal{H}_4$  in [4]. The second part is a straightforward consequence of our Hypothesis  $\mathcal{H}_2$ .

The aim is now to propose a methodology to get a pointwise estimate of the diffusion term. For that purpose, we first consider the drift estimation. Recall that  $g(x, y) = -[\sigma^2(x)y + \nabla V(x)]$ . We define the Nadaraya-Watson estimator :

$$H_n(x, y) = \frac{1}{(n-1)b_{1,n}^d b_{2,n}^d} \sum_{i=l_n+1}^{l_n+n-1} K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right) \frac{Y_{(i+1)h_n} - Y_{ih_n}}{h_n}. \quad (4.20)$$

We get Theorem 4.3 stated below.

**THEOREM 4.3** *Assume that Hypotheses  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are satisfied. Moreover we assume that*

*(i)  $b_{1,n}$ ,  $b_{2,n}$  and  $h_n \rightarrow 0$ ,*

$$(ii') \quad n h_n b_{1,n}^d b_{2,n}^d \rightarrow +\infty,$$

$$(iii) \quad \exists p > 1 \text{ such that } \frac{h_n}{(b_{1,n} b_{2,n})^{d(2-\frac{1}{p})}} \xrightarrow{n \rightarrow +\infty} 0,$$

$$(iv) \quad \exists m \in \mathbb{N}^* \text{ such that } n b_{1,n}^d b_{2,n}^d \max(b_{1,n}, b_{2,n})^{2(m+1)} \rightarrow 0,$$

$$(v) \quad l_n h_n \rightarrow +\infty$$

and:

•  $\mathcal{H}_3$   $g$  belongs to the domain of the infinitesimal generator  $\mathcal{L}$ , in all  $L^\gamma(\mu)$ , for  $1 \leq \gamma < +\infty$ .

Then, starting from any initial condition  $z_0 = (x_0, y_0)$ , we get :

$$\mathbb{E} \left( \frac{H_n}{\tilde{p}_s^\beta}(x, y) - g(x, y) \right)^2 = O \left( \frac{1}{n b_{1,n}^d b_{2,n}^d h_n} \right). \quad (4.21)$$

*Remark 4.4* From the fluctuation-dissipation relation, we have an analytical formula for the stationary density  $p_s^\beta$ . From this formula, we can prove that Hypothesis  $\mathcal{H}_3$  is satisfied as soon as the potential  $V$  has a polynomial growth as  $x$  tends to infinity.

*Proof.* In the following, we consider the stationary case, that is we start from the stationary distribution:  $(X_0, Y_0) \sim p_s^\beta(z)dz$  with  $p_s^\beta(z)$  described by (2.4), and  $l_n \equiv 0$ :

$$H_n(x, y) = \frac{1}{(n-1)b_{1,n}^d b_{2,n}^d} \sum_{i=1}^{n-1} K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right) \frac{Y_{(i+1)h_n} - Y_{ih_n}}{h_n}.$$

The case where we start from any initial point  $z_0 = (x_0, y_0)$  can then be deduced from the stationary case, following the proof in [2, Theorem 5.1].

Step 1: Let us first prove that

$$\mathbb{E}[H_n(x, y)] \rightarrow g(x, y) p_s^\beta(x, y) \text{ as } n \text{ tends to infinity.} \quad (4.22)$$

We define the auxiliary sequence  $(n-1)b_{1,n}^d b_{2,n}^d \hat{H}_n(x, y)$  as:

$$\sum_{i=1}^{n-1} K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right) \frac{\sigma(X_{ih_n})(W_{(i+1)h_n} - W_{ih_n}) + \int_{ih_n}^{(i+1)h_n} g(X_t, Y_t) dt}{h_n}.$$

Then  $(n-1)h_n b_{1,n}^d b_{2,n}^d (H_n(x, y) - \hat{H}_n(x, y))$  is equal to:

$$\sum_{i=1}^{n-1} K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right) \int_{ih_n}^{(i+1)h_n} [\sigma(X_t) - \sigma(X_{ih_n})] dW_t. \quad (4.23)$$

Let  $\|\cdot\|$  denote the euclidean norm in  $\mathbb{R}^d$ . Then, applying the Itô isometry to (4.23), we bound

$h_n^2(n-1)^2 b_{1,n}^{2d} b_{2,n}^{2d} \mathbb{E} \|H_n(x, y) - \hat{H}_n(x, y)\|^2$  by:

$$\sum_{i=1}^{n-1} \mathbb{E} \left( K^2 \left( \frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}} \right) \int_{ih_n}^{(i+1)h_n} \text{Tra}((\sigma(X_t) - \sigma(X_{ih_n}))^2) dt \right). \quad (4.24)$$

Thus, as the first derivative of  $\sigma$  is uniformly bounded, we deduce from (4.24) and from a Taylor formula at order one that:

$$\mathbb{E} \|H_n(x, y) - \hat{H}_n(x, y)\|^2 = O \left( \frac{h_n}{(n-1)b_{1,n}^d b_{2,n}^d} \right). \quad (4.25)$$

We then define  $(n-1)b_{1,n}^d b_{2,n}^d \tilde{H}_n(x, y)$  as

$$\sum_{i=1}^{n-1} K \left( \frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}} \right) \left( \frac{\sigma(X_{ih_n})(W_{(i+1)h_n} - W_{ih_n})}{h_n} + g(X_{ih_n}, Y_{ih_n}) \right).$$

Then  $(n-1)h_n b_{1,n}^d b_{2,n}^d (\hat{H}_n(x, y) - \tilde{H}_n(x, y))$  is equal to:

$$\sum_{i=1}^{n-1} K \left( \frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}} \right) \int_{ih_n}^{(i+1)h_n} (g(X_t, Y_t) dt - g(X_{ih_n}, Y_{ih_n})) dt.$$

Let  $P_t$  denote the semigroup associated to the Markov process  $Z_t$ . Using the super index for denoting the coordinates of a vector, we get:

$$\begin{aligned} & \mathbb{E} \left( \hat{H}_n^k(x, y) - \tilde{H}_n^k(x, y) \right) \\ &= \frac{1}{b_{1,n}^d b_{2,n}^d} \int K \left( \frac{x-u}{b_{1,n}}, \frac{y-v}{b_{2,n}} \right) \frac{1}{h_n} \int_0^{h_n} \left( P_t(g^k)(u, v) - g^k(u, v) \right) dt p_s^\beta(u, v) du dv \\ &= \int K(u_1, v_1) du_1 dv_1 \times \\ & \quad \int_0^{h_n} \frac{t}{h_n} \left( \frac{(P_t - I)(g^k)}{t}(x - b_{1,n}u_1, y - b_{2,n}v_1) \right) dt p_s^\beta(x - b_{1,n}u_1, y - b_{2,n}v_1) = o(h_n). \end{aligned} \quad (4.26)$$

Indeed, from the regularity assumptions  $\mathcal{H}_0$  and  $\mathcal{H}_1$ , each coordinate of  $g$  is in the domain of  $\mathcal{L}$ , and then the dominated convergence Theorem allows to take the limit  $n \rightarrow +\infty$  in the integral. Moreover, we have:

$$\begin{aligned} & \mathbb{E}[\tilde{H}_n(x, y)] \\ &= \int K(u_1, v_1) g(x - b_{1,n}u_1, y - b_{2,n}v_1) p_s^\beta(x - b_{1,n}u_1, y - b_{2,n}v_1) du_1 dv_1 \end{aligned} \quad (4.27)$$

which tends to  $g(x, y) p_s^\beta(x, y)$  as  $n$  tends to infinity. Then, from (4.25), (4.26) and (4.27) we get (4.22).

Step 2: This step is devoted to the study of  $\mathbb{E} \left( \hat{H}_n(x, y) - \tilde{H}_n(x, y) \right)^2$ . We define:

$$\mathcal{B}_n(i) = \frac{1}{h_n} \int_{ih_n}^{(i+1)h_n} (g(X_t, Y_t) dt - g(X_{ih_n}, Y_{ih_n})) dt.$$

We have:

$$\hat{H}_n(x, y) - \tilde{H}_n(x, y) = \frac{1}{(n-1)b_{1,n}^d b_{2,n}^d} \sum_{i=1}^{n-1} K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right) \mathcal{B}_n(i).$$

The above equality can be written coordinatewise as:

$$\hat{H}_n^k(x, y) - \tilde{H}_n^k(x, y) = \frac{1}{(n-1)b_{1,n}^d b_{2,n}^d} \sum_{i=1}^{n-1} K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right) \mathcal{B}_n^k(i).$$

We have:

$$\begin{aligned} \text{Var} \left( (n-1)b_{1,n}^d b_{2,n}^d (\hat{H}_n^k(x, y) - \tilde{H}_n^k(x, y)) \right) &= \sum_{i=1}^{n-1} \text{Var} \left( K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right) \mathcal{B}_n^k(i) \right) \\ &+ \sum_{1 \leq i \neq l \leq n-1} \text{Cov} \left( K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right) \mathcal{B}_n^k(i), K\left(\frac{x - X_{lh_n}}{b_{1,n}}, \frac{y - Y_{lh_n}}{b_{2,n}}\right) \mathcal{B}_n^k(l) \right). \end{aligned}$$

To bound the first term we need to bound the second moment and the square of the first moment. As the procedure is detailed in [2] we only detail the computations of the bound for the second moment. By stationarity we have:

$$\begin{aligned} \sum_{i=1}^{n-1} \mathbb{E} \left( \left( K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right) \mathcal{B}_n^k(i) \right)^2 \right) &= (n-1) \mathbb{E} \left[ \left( K\left(\frac{x - X_0}{b_{1,n}}, \frac{y - Y_0}{b_{2,n}}\right) \mathcal{B}_n^k(0) \right)^2 \right] \\ &\leq (n-1) \frac{1}{h_n^2} \int_0^{h_n} t^2 \int K^2\left(\frac{x-u}{b_{1,n}}, \frac{y-v}{b_{2,n}}\right) \left( \frac{P_t(g^k)(u, v) - g^k(u, v)}{t} \right)^2 p_s^\beta(u, v) du dv dt \\ &= O \left( (n-1) h_n b_{1,n}^{d/p} b_{2,n}^{d/p} \right). \end{aligned} \tag{4.28}$$

This latter bound has been obtained by applying Hölder's inequality with conjugate exponents  $p$  and  $q$ . Then we used that  $\left( \frac{P_t(g^k)(u, v) - g^k(u, v)}{t} \right)$  is bounded in  $\mathbb{L}^{2q}(p_s^\beta(u, v) du dv)$  by assumption.

To bound the covariance terms, we use the covariance inequality in [5, Section 1.2.2, p. 9, Theorem 3, Item (1)], which can be applied here as the process  $(Z_t)$  is  $\beta$ -mixing (see (2.5)). In the following,  $C$  denotes some positive constant which may vary from line to line. Let  $r, p \in \mathbb{N}^*$ . We get:

$$\begin{aligned}
& \sum_{i \neq l} \text{Cov}\left(K\left(\frac{x - X_{ih_n}}{b_{1,n}}, \frac{y - Y_{ih_n}}{b_{2,n}}\right) \mathcal{B}_n^k(i), K\left(\frac{x - X_{lh_n}}{b_{1,n}}, \frac{y - Y_{lh_n}}{b_{2,n}}\right) \mathcal{B}_n^k(l)\right) \\
& \leq Cn \sum_{j=1}^{n-2} \min\left(\left(b_{1,n}^d b_{2,n}^d\right)^{1-\frac{1}{r}} e^{-\frac{ih_n \lambda}{r}}, \text{Var}\left(K\left(\frac{x - X_0}{b_{1,n}}, \frac{y - Y_0}{b_{2,n}}\right) \mathcal{B}_n^k(0)\right)\right) \\
& \leq Cn \sum_{j=1}^{n-2} \min\left(\left(b_{1,n}^d b_{2,n}^d\right)^{1-\frac{1}{r}} e^{-\frac{ih_n \lambda}{r}}, \left(b_{1,n}^d b_{2,n}^d\right)^{\frac{1}{p}} h_n\right).
\end{aligned}$$

If we choose  $r, p$  such that  $\frac{1}{r} + \frac{1}{p} = 1$ , then the right hand term in the above inequality is  $O\left(n(b_{1,n} b_{2,n})^{\frac{d}{p}}\right)$ .

Summing the variance and the covariance terms we obtain:

$$\text{Var}(\hat{H}_n^k(x, y) - \tilde{H}_n^k(x, y)) = O\left(n^{-1}(b_{1,n} b_{2,n})^{\frac{d}{p}-2d}\right). \quad (4.29)$$

Using Cauchy-Schwarz Inequality and (4.28), it is possible to prove that

$$\left(\mathbb{E} \hat{H}_n^k(x, y) - \mathbb{E} \tilde{H}_n^k(x, y)\right)^2 = O\left(h_n b_{1,n}^{\frac{d}{p}-2d} b_{2,n}^{\frac{d}{p}-2d}\right). \quad (4.30)$$

Thus, we deduce from (4.29) and (4.30),

$$\mathbb{E}\left(\hat{H}_n^k(x, y) - \tilde{H}_n^k(x, y)\right)^2 = O\left(h_n(b_{1,n} b_{2,n})^{\frac{d}{p}-2d}\right). \quad (4.31)$$

Step 3: In this last step, we want to prove that  $\mathbb{E}(H_n(x, y) - g(x, y)p_s^\beta(x, y))^2$  tends to zero as  $n$  tends to infinity. We write:

$$\mathbb{E}(H_n^k(x, y) - g^k(x, y)p_s^\beta(x, y))^2 \leq 3\left(\mathcal{J}_{1n}^k + \mathcal{J}_{2n}^k + \mathcal{J}_{3n}^k\right) \quad (4.32)$$

with  $\mathcal{J}_{1n}^k = \mathbb{E}\left(H_n^k(x, y) - \hat{H}_n^k(x, y)\right)^2$ ,  $\mathcal{J}_{2n}^k = \mathbb{E}\left(\hat{H}_n^k(x, y) - \tilde{H}_n^k(x, y)\right)^2$  and  $\mathcal{J}_{3n}^k = \mathbb{E}\left(\tilde{H}_n^k(x, y) - g^k(x, y)p_s^\beta(x, y)\right)^2$ .

From (4.25) and (4.31) we get:

$$\mathcal{J}_{1n}^k + \mathcal{J}_{2n}^k = O\left(\frac{h_n}{(n-1)b_{1,n}^d b_{2,n}^d} + h_n(b_{1,n} b_{2,n})^{\frac{d}{p}-2d}\right). \quad (4.33)$$

Let us now consider the third term,  $\mathcal{J}_{3n}^k$ , in the decomposition (4.32). The stationarity yields:

$$\mathbb{E} \tilde{H}_n^k(x, y) = \int K(u_1, v_1) g^k(x - b_{1,n}u_1, y - b_{2,n}v_1) p_s^\beta(x - b_{1,n}u_1, y - b_{2,n}v_1) du_1 dv_1.$$

The asymptotic behavior of this term can be classically handled with Hypothesis (iv)

(see e.g., [2, page 14]). It yields:

$$(n-1)b_{1,n}^d b_{2,n}^d h_n \mathcal{J}_{3n}^k = (n-1)b_{1,n}^d b_{2,n}^d h_n (\mathbb{E}[\tilde{H}_n^k(x, y) - g^k(x, y)p_s^\beta(x, y)])^2 = o(1). \quad (4.34)$$

From (4.33) and (4.34) we get for each  $k$ :

$$\mathbb{E}(H_n^k(x, y) - g^k(x, y)p_s^\beta(x, y))^2 = O\left(\frac{1}{(n-1)b_{1,n}^d b_{2,n}^d h_n} + h_n(b_{1,n}b_{2,n})^{\frac{d}{p}-2d}\right). \quad (4.35)$$

As (i) – (ii'), (iv) – (v) in Theorem 4.3 imply (i) – (ii), (iv) – (v) in Corollary 4.1, we get the bound stated in (4.19). Thus, from Slutsky's Theorem and from (4.35) we get :

$$\mathbb{E}\left(\frac{H_n}{\tilde{p}_s^\beta}(x, y) - g(x, y)\right)^2 = O\left(\frac{1}{nb_{1,n}^d b_{2,n}^d h_n} + h_n(b_{1,n}b_{2,n})^{\frac{d}{p}-2d}\right).$$

Then, choosing  $p > 1$  such that (iii) in Theorem 4.3 is satisfied, we get (4.21) which achieves the proof of Theorem 4.3.  $\blacksquare$

**COROLLARY 4.5** *Let  $(\mathbf{e}_i)_{i=1,\dots,d}$  denote the canonical basis of  $\mathbb{R}^d$ . Under Assumptions of Theorem 4.3 we have :*

$$- < \hat{g}_n(x, \mathbf{e}_i) - \hat{g}_n(x, 0), \mathbf{e}_j > \xrightarrow{\mathbb{P}} \sigma_{ij}^2(x).$$

*Proof.* It is a straightforward consequence of Theorem 4.3.  $\blacksquare$

Let us now consider an estimation procedure for  $-\beta \nabla V(x)$ . We know from (2.4) that  $p_s^\beta(x, y) = \mathbf{C}(\beta)e^{-\beta(|y|^2+V(x))}$ , thus  $\frac{\nabla_x p_s^\beta}{p_s^\beta}(x, y) = -\beta \nabla V(x)$ . Theorem 4.6 below stated a central limit theorem for  $\frac{\nabla_x \tilde{p}_s^\beta}{\tilde{p}_s^\beta}(x, y)$ .

**THEOREM 4.6** *Assume Hypotheses  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are fulfilled. Assume that the bandwidths  $b_{1,n}$ ,  $b_{2,n}$  and the discretization step  $h_n$  satisfy Hypotheses (i), (ii), (iii), (iv) and (v) of Corollary 4.1. Assume that the kernel  $K$  is twice continuously differentiable on its bounded support.*

*Then, starting from any initial condition  $z_0 = (x_0, y_0)$ , we get :*

$$\sqrt{nb_{1,n}^{(d+2)}b_{2,n}^d} \left( \frac{\nabla_x \tilde{p}_s^\beta}{\tilde{p}_s^\beta}(x, y) + \beta \nabla V(x) \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \mathbf{L}(x, y)),$$

where  $\mathbf{L} = (l_{i,j}(x, y))_{1 \leq i, j \leq d}$  is a diagonal square matrix with

$$l_{jj}(x, y) = \frac{1}{p_s^\beta(x, y)} \int (\partial_{x_j} K(u, v))^2 du dv,$$

where  $p_s^\beta(x, y)$  is described by (2.4).

*Proof.* In the following, we consider once more the stationary case, that is we start from the stationary distribution:  $(X_0, Y_0) \sim p_s^\beta(z)dz$  with  $p_s^\beta$  described by (2.4) and  $l_n \equiv 0$ .

The extension to the non-stationary case is classical and is left to the reader. We have :

$$\begin{aligned}
\mathcal{A}_n(x, y) &:= \left( \frac{\nabla_x \tilde{p}_s^\beta}{\tilde{p}_s^\beta}(x, y) + \beta \nabla V(x) \right) \\
&= \left( \frac{\nabla_x \tilde{p}_s^\beta}{\tilde{p}_s^\beta}(x, y) - \frac{\nabla_x p_s^\beta}{\tilde{p}_s^\beta}(x, y) \right) + \left( \frac{\nabla_x p_s^\beta}{\tilde{p}_s^\beta}(x, y) + \beta \nabla V(x) \right) \\
&= \frac{1}{\tilde{p}_s^\beta(x, y)} (\nabla_x \tilde{p}_s^\beta(x, y) - \nabla_x p_s^\beta(x, y)) - \frac{\nabla_x p_s^\beta}{\tilde{p}_s^\beta p_s^\beta}(x, y) (\tilde{p}_s^\beta(x, y) - p_s^\beta(x, y)).
\end{aligned}$$

Then, as in the proof of Theorem 3.3 in [1], we can prove that the second term in the right hand side of the last equality above is  $O_{\mathbb{P}_z}((nb_{1,n}^d b_{1,n}^d)^{-1/2})$ . Recaling that  $\mathcal{D}$  denotes the convergence in distributions of probability measures, we have

$$\begin{aligned}
\mathcal{D} \lim_{n \rightarrow \infty} \sqrt{nb_{1,n}^{(d+2)} b_{2,n}^d} \mathcal{A}_n(x, y) \\
&= \mathcal{D} \lim_{n \rightarrow \infty} \sqrt{nb_{1,n}^{(d+2)} b_{2,n}^d} \frac{1}{\tilde{p}_s^\beta(x, y)} (\nabla_x \tilde{p}_s^\beta(x, y) - \nabla_x p_s^\beta(x, y)) \\
&= \frac{1}{p_s^\beta(x, y)} \mathcal{D} \lim_{n \rightarrow \infty} \sqrt{nb_{1,n}^{(d+2)} b_{2,n}^d} (\nabla_x \tilde{p}_s^\beta(x, y) - \nabla_x p_s^\beta(x, y)),
\end{aligned}$$

where the last line was obtained from Slutsky's Theorem.

In the following we sketch the study of the convergence in distribution of  $\mathcal{R}_n := \sqrt{nb_{1,n}^{(d+2)} b_{2,n}^d} (\nabla_x \tilde{p}_s(x, y) - \nabla_x p_s^\beta(x, y))$ . Let  $\partial x_l$  denote the partial derivative with respect to the  $x_l$ -variable. We get from the covariance inequality in [5, Section 1.2.2, p. 9, Theorem 3, Item (1)] :

$$\text{Cov}(\partial_{x_j} \tilde{p}(x, y), \partial_{x_l} \tilde{p}(x, y)) = O\left(\frac{p_s^\beta(x, y)}{nb_{1,n}^{(d+2)} b_{2,n}^d} \int \partial_{x_j} K(u, v) \partial_{x_l} K(u, v) du dv\right).$$

Hence and as a consequence that the kernel  $K$  has bounded support we have

$$(nb_{1,n}^{(d+2)} b_{2,n}^d) \text{Cov}(\partial_{x_j} \tilde{p}(x, y), \partial_{x_l} \tilde{p}(x, y)) \rightarrow \delta_{jl} p_s^\beta(x, y) \int (\partial_{x_j} K(u, v))^2 du dv,$$

where  $\delta_{ij}$  stand for the Kronecker symbol.

The random sequence  $\mathcal{R}_n$  is a sum of a triangular arrays of  $\beta$ -mixing random vectors in  $\mathbb{R}^d$ . An easy modification via the Cramer-Wald device allows to extend [1, Theorem 4.3] to random vectors. Then, defining  $\mathbf{D}(x, y) = (d_{ij}(x, y))$  as a diagonal matrix and if the sequences  $h_n$ ,  $b_{1,n}$  and  $b_{2,n}$  satisfy Hypotheses (i), (ii)', (iii)' and (iv), we get

$$\mathcal{R}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{D}(x, y)) \text{ where } d_{jj}(x, y) = p_s^\beta(x, y) \int (\partial_{x_j} K(u, v))^2 du dv.$$

It thus yields :

$$\sqrt{nb_{1,n}^{(d+2)}b_{2,n}^d}\mathcal{A}_n(x,y)\xrightarrow{\mathcal{D}}N(0,\mathbf{L}(x,y))$$

where  $\mathbf{L}$  is also diagonal and  $l_{jj}(x,y) = \frac{1}{p_s^\beta(x,y)} \int (\partial_{x_j} K(u,v))^2 dudv$ . ■

## 5. Numerical experiments

This section is devoted to numerical experiments on inference issues for System (1.2). For the simulations, we chose  $l_n \equiv 0$  in Formula (4.18).

### 5.1. A constant diffusion term

We first consider the particular case where  $V(x) = 6x^6 - 12x^4 + 6x^2 + \frac{1}{7}$ ,  $s(x) \equiv 2$  and  $\beta = 2$ . This particular case enters the framework of Section 3, Item 1. We thus consider the following system:

$$\begin{cases} dX_t = Y_t dt \\ dY_t = 2dW_t - (4Y_t + 36X_t^5 - 48X_t^3 + 12X_t)dt. \end{cases} \quad (5.36)$$

We will first implement the nonparametric estimation of the invariant density

$$p_s(x,y) = C \exp\left(-y^2 - 12x^6 + 24x^4 - 12x^2 - \frac{2}{7}\right)$$

and of the drift term

$$g(x,y) = -4y - 36x^5 + 48x^3 - 12x.$$

In the following, we make use of the explicit Euler-Verlet-Midpoint scheme (see e.g., [8, (2.50) p. 94]) in which the deterministic Verlet step has been corrected using a generalized Metropolis-Hastings strategy (see [8, Algorithm 2.11, p. 95]) to simulate an approximated discrete sampling  $(\tilde{X}_i, \tilde{Y}_i)_{i \in \mathbb{N}}$  of  $(X_t, Y_t)_{t \in \mathbb{R}_+}$ .

#### Estimation of the invariant density

We estimate the invariant density  $p_s^\beta$  using the kernel estimator defined by (4.18) in Section 4, and by replacing  $(X_i, Y_i)_{i \in \mathbb{N}}$  by  $(\tilde{X}_i, \tilde{Y}_i)_{i \in \mathbb{N}}$ . The invariant density is estimated on a grid  $(z_l)_{l=1,\dots,L} = (x_l, y_l)_{l=1,\dots,L}$  of size  $L = 55 \times 39$  of  $\left[\min\{\tilde{X}_i - b_{1,n}\}_{i=1}^n, \max\{\tilde{X}_i + b_{1,n}\}_{i=1}^n\right] \times \left[\min\{\tilde{Y}_i - b_{2,n}\}_{i=1}^n, \max\{\tilde{Y}_i + b_{2,n}\}_{i=1}^n\right]$ .

We use the Epanechnikov kernel defined by  $K(u,v) = K_e(u) \times K_e(v)$  with  $K_e(w) = \frac{3}{4}(1-w^2)1_{|w| \leq 1}$ , and we consider  $h_n = n^{-\gamma}$ ,  $b_{1,n} = n^{-\alpha_1}$  and  $b_{2,n} = n^{-\alpha_2}$ , with  $\alpha_1 = \alpha_2 = 0.2$  and  $\gamma = 0.15$ .

The relative mean integrated squared error below is computed on  $M = 30$  samples,



each of size  $n$ ,

$$\frac{1}{M} \sum_{m=1}^M \frac{\frac{1}{L} \sum_{l=1}^L \left( \tilde{p}_s^{(m)}(z_l) - p_s(z_l) \right)^2}{\frac{1}{L} \sum_{l=1}^L (p_s(z_l))^2}, \text{ and we obtain:}$$

$n$	$10^3$	$10^4$	$10^5$	$10^6$
Relative mean integrated squared error	0.2225	0,03841	0,0071	0,0012

As expected, the relative error decreases as  $n$  increases.

For a sample size of  $n = 10^6$  we obtain the following graphics: the discretized theoretical bivariate invariant density and its estimated version are drawn in Figure 1. The estimation of the marginal invariant density for the position (*resp.* velocity) is drawn in the left (*resp.* right) hand part of Figure 2.

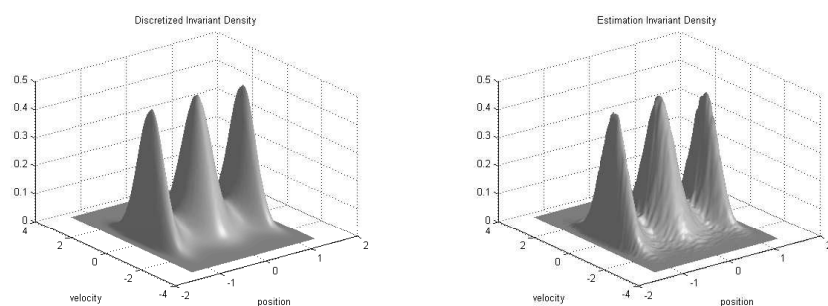


Figure 1. Discretized theoretical (left) and estimated (right) invariant bivariate density for  $V(x) = 6x^6 - 12x^4 + 6x^2 + 1/7$ ,  $\sigma(x) \equiv 2$ .

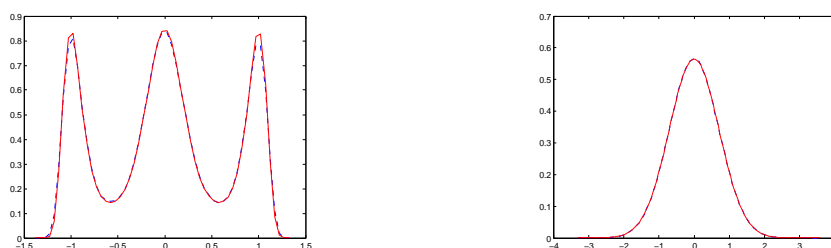


Figure 2. Theoretical (plain line) and estimated (dashed line) invariant density of the position (left) and of the velocity (right) for the potential  $V(x) = 6x^6 - 12x^4 + 6x^2 + 1/7$ .

## Estimation of the drift term

We use a regularised version of the estimator considered in Theorem 4.3:

$$\tilde{g}_n(x, y) = \frac{H_n(x, y)}{\tilde{p}_s(x, y) + \varepsilon} \quad (5.37)$$

where  $\tilde{p}_s$  has been defined in (4.18), and  $H_n(x, y)$  in (4.20), and where  $\varepsilon > 0$  is a small term whose aim is the regularisation of the estimate. We then replace, as previously,  $(X_i, Y_i)_{i \in \mathbb{N}}$  by  $(\tilde{X}_i, \tilde{Y}_i)_{i \in \mathbb{N}}$ .

The estimation of the drift in Figure 3 is drawn on a discretized grid of  $(x, y) \in [-1.15, 1.15] \times [-1.7, 1.7]$ . Note that  $[-1.15, 1.15] \times [-1.7, 1.7]$  on which the estimated invariant density of the position and the one of the velocity are both positive.

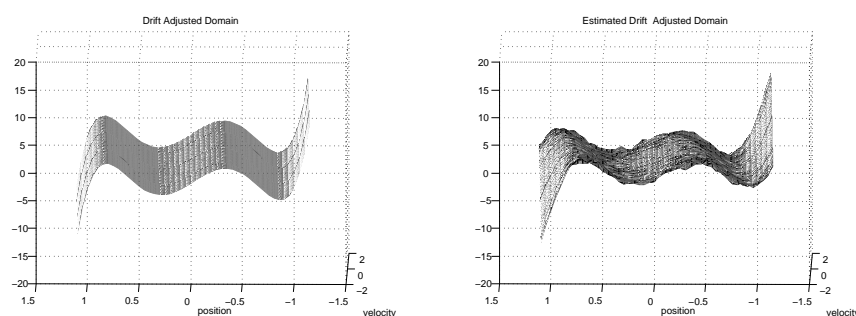


Figure 3. Discretized theoretical (left) and estimated (right) drift on  $[-1.15, 1.15] \times [-1.7, 1.7]$ .

We also provide in Figure 4 a plot of the absolute difference between the drift and its estimate. Figure 4 emphasizes the limit of the approach, as we can see some boundary effects in the estimation.

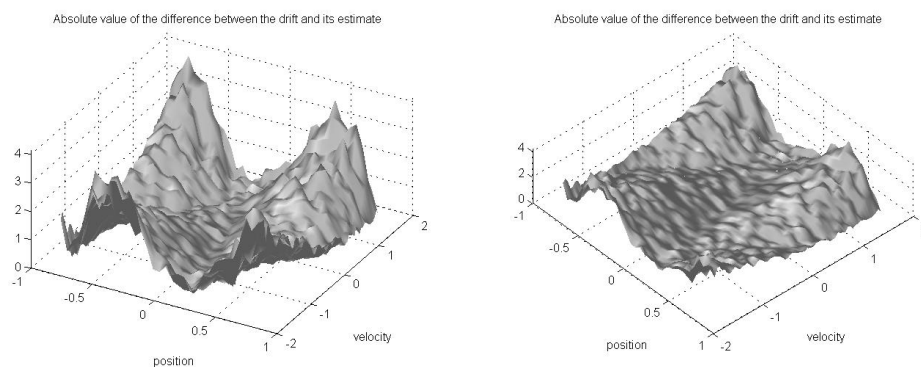


Figure 4. Absolute value of the difference between the drift and its estimate at left on the restricted domain  $[-0.85, 0.85] \times [-1.7, 1.7]$ , and a same graphic with a different orientation at right

### Estimation of the derivative of the potential

As already stated, we choose in this section the potential  $V(x) = 6x^6 - 12x^4 + 6x^2 + \frac{1}{7}$ , we

thus have:  $\nabla V(x) = 36x^5 - 48x^3 + 12x$ . We now consider the estimation of  $-\beta \nabla V(x) = -2\nabla V(x)$  as in Theorem 4.6.

On Figure 5, we present in the left hand side a discretization of  $\nabla_x p_s(x, y)$ , and in the right hand side its estimate  $\nabla_x \tilde{p}_s(x, y)$ . On Figure 6, we present in the left hand side a discretization of  $-2V'(x)$  for  $y_0 = -0.265$ , and in the right hand side its estimate  $\frac{\nabla_x \tilde{p}_s(x, y)}{\tilde{p}_s(x, y)}$ . We choose here a gaussian kernel.

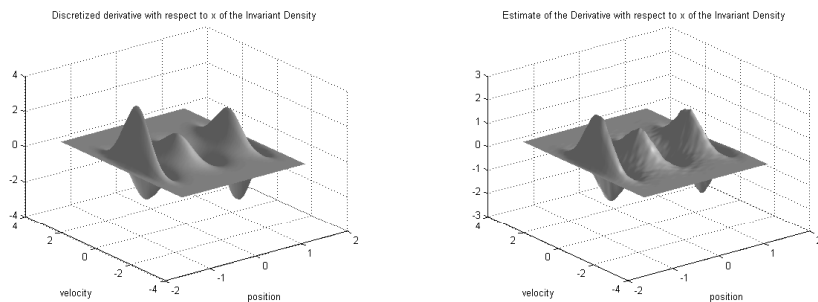


Figure 5. Discretization of  $\nabla_x p_s(x, y)$  (left), and its estimate (right).

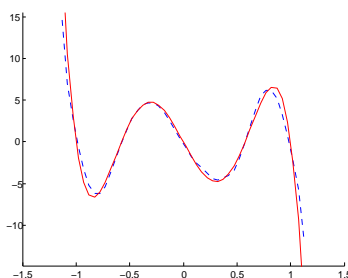


Figure 6. Discretization of the  $-2V'(x)$  (plain line) and its estimate (dashed line) for the fixed value  $y = -0.265$ .

The plot of the absolute difference for the estimation of the derivative of the potential in Figure 7 also shows some boundary effects. The correction of boundary effects could be the next aim in this framework.

Finally, we have drawn in Figure 8 below  $\tilde{g}_n(x, y) - \tilde{g}_n(x, 0)$  for two different values of  $y$ , to illustrate the result of Corollary 4.5.

We conclude that the quality of  $\tilde{g}_n(x, y) - \tilde{g}_n(x, 0)$  to estimate  $\beta^2$  depends on the value of  $y$ . It would be interesting to investigate further the asymptotic properties of this estimator, in the more general case where the diffusion term  $\sigma(x, y)$  depends on the position  $x$ .

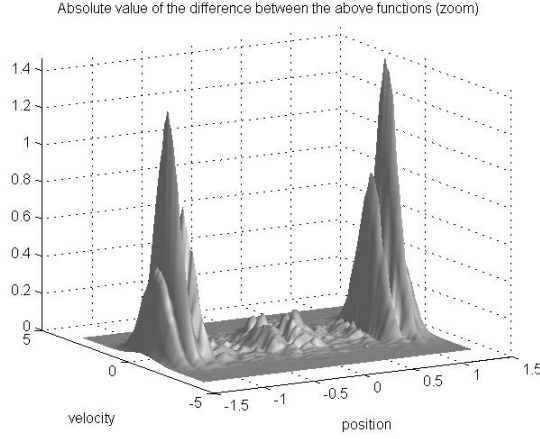


Figure 7. Absolute value of the difference between  $\nabla_x p_s$  and its estimate.

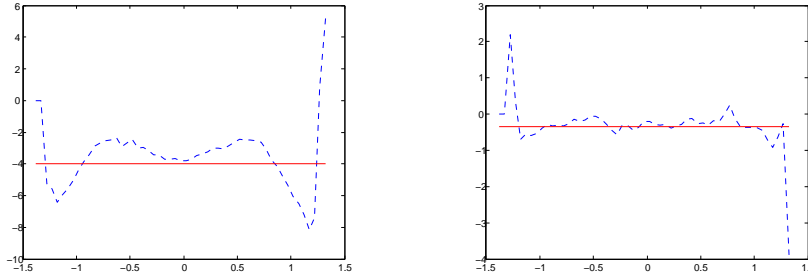


Figure 8. The dashed line corresponds to  $\tilde{g}_n(x, y) - \tilde{g}_n(x, 0)$  for  $y = 1$  (left) and for  $y = 0.085$  (right), the red line corresponds to the corresponding value  $\beta^2 \times y = 4y$ .

## 5.2. A diffusion term that depends on the position

In this section, we consider the following system:

$$\begin{cases} dX_t = Y_t dt \\ dY_t = \exp\left(-\frac{1}{X_t^2 + 1}\right) dW_t - \left(\exp\left(-\frac{2}{X_t^2 + 1}\right) Y_t + 36X_t^5 - 48X_t^3 + 12X_t\right) dt. \end{cases} \quad (5.38)$$

We will first implement the nonparametric estimation of the invariant density

$$p_s(x, y) = C \exp\left(-y^2 - 12x^6 + 24x^4 - 12x^2 - \frac{2}{7}\right),$$

which is unchanged with respect to Section 5.1, and of the drift term

$$g(x, y) = -\exp\left(-\frac{1}{x^2 + 1}\right) y - 36x^5 + 48x^3 - 12x.$$

The choice of the kernel and of the schemes for simulation are similar to the ones in Section 5.1 and are thus not recalled here.

**Estimation of the invariant density** For a sample size of  $n = 10^6$  we obtain the following graphics: the discretized theoretical bivariate invariant density and its estimated version are drawn in Figure 9.

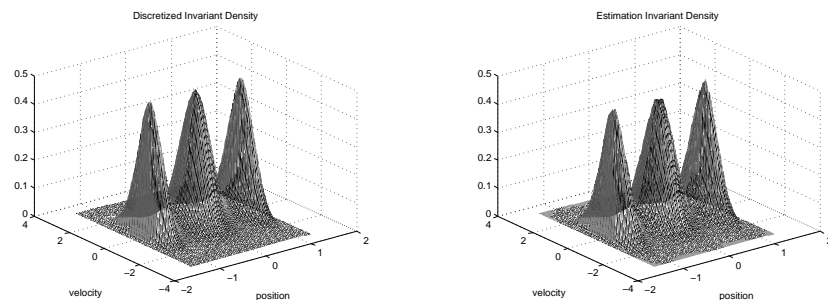


Figure 9. Discretized theoretical (left) and estimated (right) invariant bivariate density for  $V(x) = 6x^6 - 12x^4 + 6x^2 + 1/7$ ,  $\sigma(x) = \exp(-\frac{1}{x^2+1})$ .

In Figure 10 is plotted the absolute difference between the invariant density and its estimate.

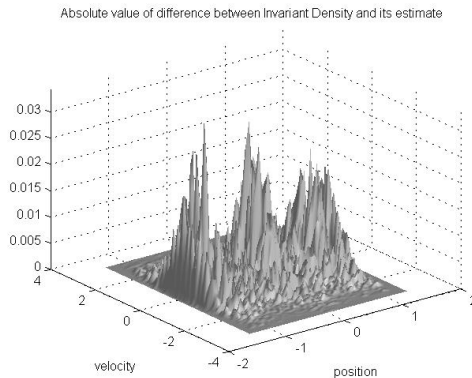


Figure 10. Absolute value of the difference between the theoretical invariant density and its estimate.

### Estimation of the drift term

The estimation of the drift in Figure 11 is drawn on a discretized grid of  $(x, y) \in [-1, 1] \times [-1.7, 1.7]$ .

Figure 12 plots the absolute difference between the drift and its estimate.

**Acknowledgments.** This work has been partially supported by the project ECOS NORD/FONACIT (France-Venezuela) **V12M01** *Equations différentielles stochastiques et analyse de sensibilité : applications environnementales*, by the 16-MATH-03 SIDRE project of the MATH-AmSud Program and by the Inria International Chairs program.

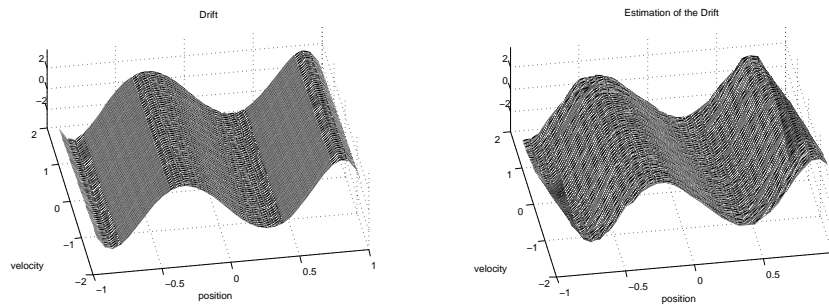


Figure 11. Discretized theoretical (left) and estimated (right) drift on  $[-1, 1] \times [-1.7, 1.7]$ .

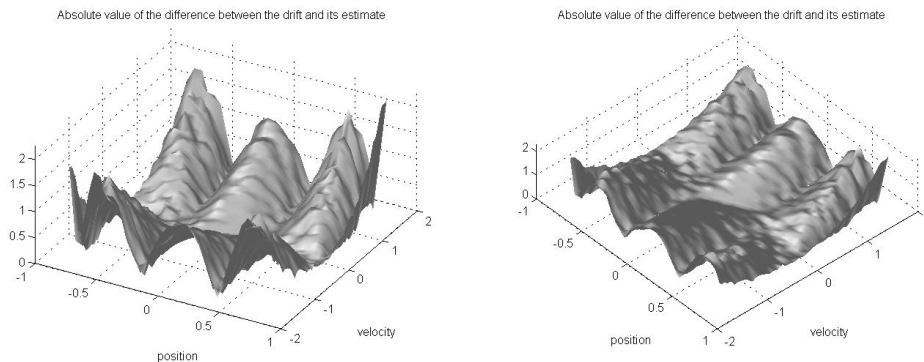


Figure 12. Absolute value of the difference between the drift and its estimate at left on the restricted domain  $[-0.85, 0.85] \times [-1.7, 1.7]$ , and the same graphic with a different orientation at right.

## References

- [1] P. Cattiaux, J. R. León and C. Prieur (2014). Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. I. Invariant density. *Stochastic Processes and their Applications* **124** (3), p. 1236-1260.
- [2] P. Cattiaux, J. R. León and C. Prieur (2014). Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. II. Drift term. *ALEA* **11** (1) p. 359-408.
- [3] P. Cattiaux, J. R. León and C. Prieur (2015). Recursive Estimation for Stochastic Damping Hamiltonian Systems. *Journal of Nonparametric Statistics* **27** (3), p. 401-424.
- [4] P. Cattiaux, J. R. León and C. Prieur (2016). Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. III. Diffusion term. *Annals of Applied Probability* **26** (3), p. 1581-1619.
- [5] P. Doukhan (1994). *Mixing. Properties and examples*. Lecture Notes in Statistics, **85**. Springer-Verlag, New York.
- [6] D. Down, S. P. Meyn and R. Tweedie (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23** (4), p. 1671-1691.
- [7] I. Horenko, and Ch. Schütte (2008). Likelihood-Based Estimation of Multidimensional Langevin Models and its Application to Biomolecular Dynamics. *Mult. Mod. Sim.* **7** (2), p. 731-773.

- [8] T. Lelièvre, M. Rousset and G. Stoltz (2010). Free energy computations: a mathematical perspective. *Imperial College Press*.
- [9] J. Mattingly and A.M. Stuart (2002) Geometric ergodicity of some hypo-elliptic diffusions for particle motions. *Markov Process. Related Fields*. **vol. 8 no. 2** , p. 199-214.
- [10] V. Konakov, S. Menozzi and S. Molchanov (2010). Explicit parametrix and local limit theorems for some degenerate diffusion processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **46 (4)**, p. 908-923.
- [11] A. Szepessy (2011). Langevin molecular dynamics derived from Ehrenfest dynamics. *Math. Models Methods Appl. Sci.* 21, no. 11, 2289-2334.
- [12] G. Viennet (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Relat. Fields*, **107**, p. 467-492.
- [13] L. Wu (2001). Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stoch. Processes and Appl.* **91**, p. 205-238.